
WikiChron networks: Una herramienta para el análisis de redes en wikis



Trabajo de Fin de Grado
Curso 2018–2019

Autor
Youssef El Faqir El Rhazoui

Director
Javier Arroyo Gallardo
Abel Serrano Juste

Grado en Ingeniería Informática
Facultad de Informática
Universidad Complutense de Madrid

WikiChron networks: Una herramienta para el análisis de redes en wikis

Trabajo de Fin de Grado en Ingeniería Informática
Departamento de Ingeniería del Software e Inteligencia
Artificial

Autor
Youssef El Faqir El Rhazoui

Director
Javier Arroyo Gallardo
Abel Serrano Juste

Convocatoria: *Junio 2019*
Calificación:

Grado en Ingeniería Informática
Facultad de Informática
Universidad Complutense de Madrid

31 de mayo de 2019

Agradecimientos

Aún siendo un desarrollo en el que estoy yo solo como integrante, este ha sido fruto gracias a la dedicación y desempeño de un magnífico equipo de profesionales.

Han sido muchas semanas de constantes reuniones y tardes muy largas con Abel Serrano, limpiando código y pensando constantemente en el siguiente paso a dar.

Mi reconocimiento a la gran labor docente de mi director, Javier Arroyo, que pese a su baja, su presencia a sido constante para impulsar este trabajo, y para introducirme en el mundo investigador, que finalmente pudimos cristalizar en un artículo en conjunto con Abel.

Quiero mencionar también a Guillermo Jiménez y a Elena Martínez, que pese a estar muy ocupados hemos podido contar con su ayuda para testear y dar diseño a la interfaz.

A todos ellos, muchas gracias.

Resumen

La producción colaborativa de bienes comunes ha marcado el desarrollo de grandes proyectos como puedan ser *Wikipedia* o *Linux*. Estos proyectos se sustentan sobre grupos de personas de diferente índole, que trabajan cooperativamente sin jerarquías marcadas. Esto hace que el análisis de redes complejas sea perfecto para abordar el problema. Pero debido a que se conforman comunidades dinámicas y constantemente cambiantes, debemos tener en cuenta como eje central del análisis la variable temporal.

En este trabajo proponemos una herramienta web, con la que facilitar el estudio y seguimiento de las comunidades que surgen y se conforman alrededor de wikis. La herramienta dispone de tres visualizaciones diferentes, en función del espacio de interacción seleccionado (artículos, páginas de talk y páginas de usuario), además de presentar métricas que resumen comportamientos y niveles de participación de los usuarios dentro de la comunidad. De esta manera, se puede tener diferentes perspectivas de la comunidad en espacios de interacción como en el plano temporal.

Por último, ilustramos el uso de *WikiChron networks* con un caso de estudio para mostrar las virtudes que pensamos que ofrece la herramienta.

Palabras clave

Wiki, Open Collaboration, Network Analysis, Complex Networks, Visualization, Webtool

Abstract

Commons-based peer production communities can be analyzed with the help of social network analysis. However, they are fluid organizations that continually change over time and consequently the time dimension needs to be considered in the analysis as a central variable.

In this work we propose a webtool, *WikiChron networks*, to facilitate the study of the evolution of wiki communities through time. The tool displays three different community networks depending on the space considered for the interactions: articles, talk pages of articles or talk pages of users. The joint consideration of these three networks offer complementary views of the same community, while the time dimension makes possible to observe how the network structures changes through time and the changes in the played network role experimented by some editors. We illustrate the usefulness of our tool analyzing the evolution of a wiki community in different moments and showing network structures that can be seen in other wiki communities.

Keywords

Wiki, Open Collaboration, Network Analysis, Complex Networks, Visualization, Webtool

Índice

1. Introduction	1
1.1. Objectives	1
1.2. Document structure	2
1. Introducción	3
1.1. Objetivos	3
1.2. Estructura del documento	4
2. Estado del arte	5
2.1. Ediciones y roles de editores	6
2.2. Distribución del trabajo en las wikis	7
2.3. Redes complejas	7
2.4. Herramientas de visualización de wikis	8
3. Definición de las wikis como redes complejas	11
3.1. Estructura de los datos en una wiki	11
3.2. Discusión de las redes elegidas	14
3.2.1. Definición de la dimensión temporal	15
3.3. Definición de la red de comunicación directa entre usuarios . .	15
3.4. Definición de la red de discusión de artículos como red de colaboración y comunicación	16
3.5. Definición de la red de Co-edición como red de colaboración .	17
3.6. Métricas que miden la importancia relativa de los usuarios . .	17
3.6.1. Métricas de producción	18
3.6.2. Centralidad	18
3.7. Métricas que muestran el comportamiento de la comunidad .	19
3.7.1. Búsqueda de comunidades	21
4. Arquitectura de la aplicación y modelo de proceso	23

4.1. Tecnologías	23
4.2. Arquitectura	25
4.2.1. Proceso ETL de los datos	26
4.3. Modelo de proceso	27
4.3.1. Primera fase	27
4.3.2. Segunda fase	28
4.4. Ventajas y limitaciones técnicas	29
5. Guía de uso	31
6. Análisis	37
6.1. El caso de Cities Wiki	37
6.1.1. Primera etapa	37
6.1.2. Segunda etapa	40
6.1.3. Conclusiones	42
7. Trabajo futuro	45
8. Conclusiones	47
8. Conclusions	49
Bibliografía	51

Índice de figuras

3.1. Histórico de revisiones de Harry Potter Wiki	12
3.2. Algunos de los principales namespaces comunes en MediaWiki	12
3.3. Página de talk asociada al artículo; "Newton Scamander" de Harry Potter	13
3.4. Una página de user talk de un usuario en la wiki de Harry Potter	13
4.1. Modelo de tres capas de las tecnologías usadas	26
4.2. Modelo de tres capas de los archivos y paquetes involucrados. Estos se encuentran en la ruta WikiChron/wikichron/dash/apps/networks	26
4.3. Primera versión de WikiChron networks	28
5.1. Página de inicio de WikiChron	31
5.2. Página de selección previa	32
5.3. Red de co-edición de la wiki Community Data Collective Wiki	33
5.4. Clusters coloreados de diferentes colores	33
5.5. Degradado de la longevidad de los usuarios. Azul oscuro indica usuarios más antiguos. El tamaño refiere a la cantidad de ediciones en artículos.	34
5.6. Selección de una fila de la tabla, e información concreta de un nodo.	35
6.1. Red de co-edición de artículos entre Nov 2006 - Nov 2007 . .	38
6.2. Red de co-edición de artículos entre Nov 2006 - Feb 2007 . .	39
6.3. Red de comunicación directa entre Nov 2006 - Feb 2007 . .	40
6.4. Red de coedición entre Nov 2010 - Feb 2011	41
6.5. Red de coedición entre Nov 2016 - Feb 2017	42
6.6. Nuevos Artículos	43
6.7. Usuarios nuevos	43

Índice de tablas

3.1. Tipos de red en base al contenido de la wiki seleccionado. . .	15
6.1. Un top de usuarios con las métricas más altas de la red de co-edición de artículos entre Nov 2006 - Feb 2007	39
6.2. Resumen de las métricas más importantes de las redes de co-edición	44

Chapter 1

Introduction

In 1995, Ward Cunningham invented the first wiki system that allowed people to share content through the World Wide Web, until then, users required more technical knowledge to be able to perform these tasks.

In 2003, Wikipedia was born and has not stopped growing since then. Nowadays, it is undoubtedly the largest online encyclopedia in the world. Although, Wikipedia has been a success, we must also highlight that other platforms such as Wikia (Fandom), that host thousands of wikis in order to producing content created by different communities of hundreds (even thousand) editors that are self-organized.

As a result, wikis are one of the most complex social media networks to analyze. Thus, we can study these communities to understand how they behave, evolve and produce collaborative content throughout their life.

Social network analysis provides us to study the relationships that occur within these communities, seeing how connections between users are produced, giving a broad sense to the underlying structure, and understanding user's dynamic more closely.

1.1. Objectives

With this project we seek to give more visibility to Social network analysis as an analytical tool to understand the meaning of wiki communities.

To do this, we have developed a tool, *WikiChron networks*, to facilitate the study and management of wikis data such as complex networks and open

systems.

The premise under *WikiChron networks* is to be easy to use and allow for a deeper analysis of how the community behaves, for this, we have always had in mind the temporal dimension, since it allows a more focused study of the problem.

In addition, we wanted to follow the philosophy of open data that wikis have, and for this, we have tried to make it as simple as possible to share and download the data that the tool hosts.

1.2. Document structure

In chapter 2, we give an introduction to the paradigm of wikis, illustrating the achievements that have been reached in their study. We also talk about other tools that are available to compare.

In the chapter; Wikis as networks (3) we give a vision of how to model the problem in the form of a graph, what alternatives can be proposed and what parameters are proposed for the analysis.

Next, chapter 4 explains how the information that is used in the tool has been extracted, what technologies it uses, and a usage guide is given.

Chapter 4.3 gives a temporary vision of how work has been done to build the final web tool. Finally, the chapter of analysis 6 poses a case of study with the tool.

Capítulo 1

Introducción

En 1995, Ward Cunningham inventó el primer sistema de wikis que permitió a la gente compartir contenido a través de la *World Wide Web*, hasta entonces los usuarios necesitaban tener conocimientos más técnicos, para poder realizar estas tareas.

En 2003, nace *Wikipedia* y no ha parado de crecer desde entonces. Hoy en día, es sin duda alguna la enciclopedia online más grande que hay. Aunque *Wikipedia* haya sido un éxito, también debemos de destacar que otras plataformas como *Wikia (Fandom)* que albergan miles de wikis, todas ellas con la finalidad de producir contenido creado por la comunidad.

Las wikis son un recurso muy valioso de información al que podemos acceder dada su naturaleza de ser un proyecto open source. Debido a esto, las wikis conforman una red social de las más complejas a las que se puede acceder. Así, podemos estudiar estas comunidades para entender como se comportan, evolucionan y desarrollan para producir contenido colaborativo a lo largo de su vida útil.

El análisis de redes complejas nos permite estudiar las relaciones que se producen dentro de estas comunidades, ver como se producen estas conexiones entre usuarios, dar un sentido amplio a la estructura subyacente de coordinación y entender de una manera más cercana las dinámicas de los usuarios.

1.1. Objetivos

Con este proyecto buscamos dar más visibilidad al análisis de redes como herramienta de análisis para entender el funcionamiento de las comunidades

de estas wikis.

Para ello, hemos desarrollado una herramienta, *WikiChron networks*, para facilitar el estudio y manejo de los datos de las wikis como redes complejas y sistemas abiertos.

La premisa de *WikiChron networks* por tanto, es ser fácil de usar y permitir realizar análisis más profundos de cómo se comporta la comunidad, para ello, hemos tenido siempre en mente la dimensión temporal, ya que permite un estudio más focalizado del problema.

Además, hemos querido seguir la filosofía de datos abiertos que nos dan las wikis, y por ello hemos intentado que sea lo más sencillo posible compartir y descargar los datos que aloja la herramienta.

1.2. Estructura del documento

En el capítulo 2, ofrecemos una introducción al paradigma de las wikis, ilustrando los hitos a los que se ha llegado en su estudio. También hacemos una recapitulación de que otras herramientas están disponibles en el mercado que sean comparables a la que nosotros proponemos.

En el capítulo; Wikis como redes (3) damos una visión de como modelar el problema en forma de grafo, que alternativas se pueden plantear y que métricas se proponen para el posterior análisis.

A continuación, el capítulo 4 explica cómo se ha extraído la información que se usa en la herramienta, que tecnologías usa, y se da una guía de uso.

El capítulo 4.3 da una visión temporal de cómo se ha trabajado hasta construir la herramienta final. Por último, el capítulo de análisis 6 plantea un caso de estudio real con la herramienta.

Capítulo 2

Estado del arte

Como ya hemos discutido previamente el trabajo colaborativo se ha convertido en un pilar fundamental de internet, el cual ha roto las barreras jerárquicas tradicionales en pro de una democratización a la hora de generar contenido.

Las wikis son una caso curioso dentro de este paradigma, ya que a priori sin necesidad de tener un conocimiento muy técnico permiten crear grandes volúmenes de contenido entre pares.

Un caso de éxito sin duda es *Wikipedia*, que a pesar de poder editar "cualquiera", es la compilación de saber más grande de la humanidad, y en general el contenido se intenta contrastar con diferentes fuentes, para que sea fiable. De hecho, varios [estudios](#) han mostrado que los artículos relacionados con varios campos científicos, como oncología o farmacéutica, son tan precisos y contrastados como los artículos científicos de la rama del conocimiento asociada.

Siendo *Wikipedia* uno de los casos de éxito más sonados podemos destacar otros tantos, como es el caso de *Fandom* la cual conforma una de las mayores plataformas para el desarrollo e implantación de wikis de cualquier temática. Casos de éxito son el de [Harry Potter Marvel](#) o [Star Trek](#), los cuales mantienen una alta participación de sus usuarios y siguen generando con el paso del tiempo una gran cantidad de contenido.

Aunque la temática de estas wikis es variada, bien es cierto que la mayoría que consigue mantener el interés de sus usuarios y generar contenido suelen ser las que se relacionan con la temática fan, como se puede ver en el censo de Jimenez-Diaz et al. (2018). Con lo cual otras muchas wikis son abandonadas por sus usuarios, o simplemente no tienen el suficiente calado para atraer a nuevos usuarios que finalmente realicen nuevas aportaciones.

Todo lo anterior en su conjunto, nos llama a profundizar en como se forman las relaciones entre usuarios que puedan cristalizar en contenido para la comunidad, o como afecta el flujo de usuarios a la longevidad del proyecto, incluso cuando podemos entender que la wiki fracasa o tiene éxito.

2.1. Ediciones y roles de editores

Un enfoque que se ha tomado para analizar comunidades en wikis ha sido desde el punto de vista del contenido que generan los editores, es decir, que roles de editores forman estas comunidades y como afectan en ellas. Estos roles se determinan en base a una clasificación previa de las ediciones realizadas.

Ya Faigley y Witte (1981) clasificaron las ediciones sobre cualquier texto en función de la intencionalidad del cambio, dando lugar a dos grandes grupos; Los cambios de *Texto Base* que engloban cambios en el contenido del corpus del texto, y los cambios de *Superficie* los cuales hacen referencia a cambios en la forma del corpus del texto (e.g. estilo, referencias, ...).

Más adelante, con *Wikipedia* ya asentada, Ang et al. (2006) buscan encontrar diferencias culturales en el proceso de escribir artículos en esta. Con lo que comparan ediciones y artículos de diferentes idiomas como el alemán, el francés o el japonés, aunque sin basarse en taxonomías anteriores, usando así una propia para el análisis. Con esto, intentan buscar contrastes entre colectivismo–individualismo o feminidad–masculinidad y evidenciar diferencias culturales.

Posteriormente, Daxenberger y Gurevych (2012) realizan un clasificador de ediciones basándose en la taxonomía que propusieron Faigley y Witte (1981), añadiendo además un nuevo tipo, *las Políticas de Wikipedia*, que hacen referencia a ediciones que se puedan considerar vandálicas o a las propias que revierten estos vandalismos.

Con esta taxonomía, posteriormente en Arazy et al. (2016) añaden una capa de complejidad más y realizan una clasificación de los roles asociados a cada tipo de edición, y al contrario de lo que piensan en un inicio (los roles son muy marcados y van asociados a cada tipo de usuario) se dan cuenta que estos roles son más bien dinámicos, aparecen y desaparecen según la comunidad necesita de ellos.

2.2. Distribución del trabajo en las wikis

Kittur y Kraut (2010) realiza un estudio estadístico de una muestra de casi 7,000 wikis (de Wikia ahora Fandom), y hace una comparación directa con el caso *Wikipedia*. Encuentra que en ambos casos hay una relación directa entre el número de usuarios y la cantidad de recursos que se destinan a la coordinación para escribir artículos. Saca a la luz también, que las wikis al contrario que *Wikipedia* no son tan rígidas con las normas de la comunidad. Por último, evidencia la desigualdad que se produce en estas wikis, ya que la mayoría de contenido es añadido por una minoría de usuarios (*core*).

Más adelante, Shaw y Hill (2014) inciden de nuevo sobre la idea de que muchas veces grupos pequeños de usuarios controlan la mayoría de contenido y espacios de moderación. Van incluso más lejos afirmando que en vez de formarse estructuras democráticas en las que la participación es equitativa, se forman por el contrario estructuras oligárquicas, aunque luego bien es cierto que encuentran ejemplos de lo contrario.

En Serrano et al. (2018) inciden sobre la idea de estudiar la desigualdad de la distribución del trabajo dentro de una wiki, para ello aplican a parámetros de participación *el coeficiente de Gini*, el cual se usa en el ámbito económico para medir la desigualdad en la población. Incluso van más allá, y aplican un coeficiente que llaman *ratio 10:90*, el cual mide cuánto más hace el 10 % con más participación sobre el resto de la comunidad, esto finalmente se traduce en una confirmación de que en muchas comunidades la participación o el contenido de la wiki pende de unos pocos.

Más reciente, Jimenez-Diaz et al. (2018) realizan un censo de wikis de Fandom, evidenciando el gran sesgo que hay, como por ejemplo; Que el 75 % de las wikis tienen una media de no más de 23 artículos, incluso el 95 % de las wikis tienen una media de 177 artículos.

También lanzan otros datos interesantes como que el 71 % de las wikis están en inglés, a pesar de no ser la lengua más hablada del mundo.

2.3. Redes complejas

Aunque bien es cierto que las anteriores aproximaciones han arrojado luz a como se comportan estas comunidades, o que sesgos se producen en ellas, no dan un punto de vista acerca de como los editores se comportan de manera individual o como se relacionan con otros. Aquí pues es donde el análisis de redes nos da un enfoque más adecuado para abordar este asunto.

En el análisis de redes complejas se modelan entidades y las relaciones

entre ellas en forma de grafo. Sobre estos modelos podemos aplicar toda la teoría de grafos. Así, realizar análisis, predicciones y réplicas de sucesos. Por ejemplo, como se enseña en Barabási y Pósfai (2016) un fallo del sistema eléctrico de EE.UU en 2003 dejó sin corriente eléctrica a millones de personas durante muchas horas. Posteriormente, se modeló la red equivalente y un análisis demostró que fue debido a una caída del sistema en cascada, ya que el nodo en caer era lo suficientemente grande como para que los adyacentes no pudieran asumir la carga extra, y estos caer a su vez arrastrados por el primero.

Como se dice en Kadushin (2012) las redes sociales se clasifican como ego-céntricas, socio-céntricas o como sistemas abiertos. Las wikis las podemos clasificar dentro de este último grupo, ya que las relaciones o límites entre usuarios no están bien definidos, lo que ocasiona que sean difíciles de analizar. Esto es tanto así porque estas relaciones son volátiles, pues nuevos usuarios se unen, otros dejan de ser activos, incluso los anónimos pueden participar. Por tanto, la dimensión temporal consideramos que toma mucha relevancia en un estudio de estas características.

Una idea sobre la que se ha intentado incidir en Bolici et al. (2016) ha sido el hecho de clasificar a las comunidades de producción colaborativa (más allá de las wikis) como comunidades estigméricas, es decir, que la producción entre pares se asemeja a como se comportan las colonias de hormigas, en las que los individuos participan de forma equitativa y saben más o menos cual es su papel en el proyecto.

Hace poco Foote et al. (2019) intentaron ver si en las etapas iniciales de la vida de una wiki se cumple esto, para ello modelaron las comunicaciones entre usuarios como redes y hicieron un estudio sobre una muestra amplia de wikis, finalmente concluyen que ese patrón no es explicado al menos con las redes de comunicación que ellos definen.

2.4. Herramientas de visualización de wikis

Antes de empezar el desarrollo de la herramienta nos documentamos y buscamos qué soluciones pudieran haber en el mercado. Por desgracia o por fortuna no encontramos muchos avances en ese sentido.

La oferta disponible suelen ser herramientas de uso general, como [Gephi](#) o [Cytoscape](#), las cuales no generan ni mucho menos las redes que buscamos, sino que son herramientas a las que se les carga una red ya calculada, y sobre esta se pueden hacer cálculos y visualizar.

Por otro lado, también dimos con el siguiente [repositorio](#), que contiene código usado en Foote et al. (2019), el inconveniente era que la definición que se proponía para las redes era diferente a las que planteamos en 3.2, con lo cual finalmente se descarta su uso de igual manera.

En cuanto a lo más cercano que había era un *TFM* desarrollado el año pasado, Villa (2018), que aunque en un principio consideramos su uso, finalmente se descartó porque se hubiera tenido que reescribir gran parte del código para integrarse con *WikiChron*.

Por último, dispusimos de la herramienta desarrollada por compañeros de la UCM (Abel Serrano, Javier Arroyo), [WikiChron-classic](#), que aunque si trabajaba específicamente con wikis se encarga de realizar un análisis más estadístico y de series temporales. Con lo cual decidimos partir del núcleo de la aplicación para construir la nueva herramienta *WikiChron-networks*.

Capítulo 3

Definición de las wikis como redes complejas

Como ya hemos mencionado en el capítulo anterior, es perfectamente posible definir wikis como redes complejas, es decir, en forma de grafo, y el principal reto es definir la topología de la red (nodos y aristas). Pero antes, vamos a dar una breve introducción a cómo se organizan los datos en las wiki, para posteriormente discutir qué representaciones vamos a querer tomar o definir.

3.1. Estructura de los datos en una wiki

Este trabajo se centra en el estudio de wikis que implementen el software de *MediaWiki*, como es el caso de Wikia/Fandom. Este software por tanto, organiza los datos en un histórico en formato xml, en el se guardan todas las ediciones realizadas sobre la wiki en cuestión (ver figura 3.1), además de guardar también quién fue el que editó, a qué artículo o contenido es dirigido, o la fecha en la que se realizó.

Por otro lado, estas ediciones se clasifican en diferentes tipos, los llamados *Namespaces*, que aglutinan ediciones del mismo contenido. Los *namespaces* pueden no coincidir de una wiki a otra, pues cada una puede implementar los suyos propios, pero sí que hay unos que son universales a todas las wikis bajo el software de *MediaWiki*. Podemos ver algunos de los principales *namespaces* en la figura 3.2.

Por ejemplo, el namespace con el valor 0 corresponde a los artículos dentro de una wiki. El que tiene valor 1 hace referencia a las llamadas *talk pages*, que son páginas en las que los usuarios discuten acerca de como llevar un artículo concreto o qué hay que respetar en dicho artículo, suelen ir asociados

For any version listed below, click on its date to view it. For more help, see [Help:Page history](#).
 (cur) = difference from current version, (prev) = difference from preceding version, m = [minor edit](#), → = section edit, ← = automatic edit summary
 (Latest | Earliest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

[Compare selected revisions](#)

▪ (cur prev)	21:31, May 18, 2019	Ironyak1-bot (Talk contribs)	m	.. (97,839 bytes) (-10) .. (UK spelling (3), redundant pipes (1), Update link (1), fix section title (2))
▪ (cur prev)	23:44, May 17, 2019	Sammim (Talk contribs)	m	.. (97,849 bytes) (0) .. (Changed protection level for "Newton Scamander": move restriction ([edit=autoconfirmed] (indefinite) [move=sysop] (indefinite)))
▪ (cur prev)	23:38, May 17, 2019	Sammim (Talk contribs)	m	.. (97,849 bytes) (0) .. (Protected "Newton Scamander": Recent target of trolling ([edit=autoconfirmed] (indefinite) [move=autoconfirmed] (indefinite)))
▪ (cur prev)	20:05, May 17, 2019	RainA (Talk contribs)	m	.. (97,849 bytes) (+19) .. (spam)
▪ (cur prev)	20:04, May 17, 2019	Kuraudosetrafu (Talk contribs)	..	(97,830 bytes) (-19) .. (VisualEditor)
▪ (cur prev)	19:36, May 17, 2019	Maester Martin (Talk contribs)	..	(97,849 bytes) (+1) .. (Undo revision 1228148 by BlackWizard99 (talk))
▪ (cur prev)	19:18, May 17, 2019	BlackWizard99 (Talk contribs)	..	(97,848 bytes) (-1) .. (VisualEditor)
▪ (cur prev)	10:42, May 17, 2019	Goku20 (Talk contribs)	..	(97,849 bytes) (+16) .. (Undo revision 1228023 by RedWizard92 (talk))
▪ (cur prev)	10:35, May 17, 2019	RedWizard92 (Talk contribs)	..	(97,833 bytes) (-16) .. (Undo revision 1228012 by Goku20 (talk))
▪ (cur prev)	10:34, May 17, 2019	Goku20 (Talk contribs)	..	(97,849 bytes) (+16) .. (Undo revision 1228003 by RedWizard92 (talk))
▪ (cur prev)	10:33, May 17, 2019	RedWizard92 (Talk contribs)	..	(97,833 bytes) (-16) .. (Undo revision 1227985 by Goku20 (talk))
▪ (cur prev)	10:30, May 17, 2019	Goku20 (Talk contribs)	..	(97,849 bytes) (+16) .. (Undo revision 1227981 by RedWizard92 (talk))
▪ (cur prev)	10:29, May 17, 2019	RedWizard92 (Talk contribs)	..	(97,833 bytes) (-16) .. (Undo revision 1227961 by Goku20 (talk))
▪ (cur prev)	10:26, May 17, 2019	Goku20 (Talk contribs)	..	(97,849 bytes) (+16) .. (Undo revision 1227959 by RedWizard92 (talk))
▪ (cur prev)	10:26, May 17, 2019	RedWizard92 (Talk contribs)	..	(97,833 bytes) (-16) .. (Undo revision 1227942 by Goku20 (talk))
▪ (cur prev)	10:23, May 17, 2019	Goku20 (Talk contribs)	..	(97,849 bytes) (+16) .. (Undo revision 1227913 by RedWizard92 (talk))
▪ (cur prev)	10:18, May 17, 2019	RedWizard92 (Talk contribs)	..	(97,833 bytes) (-16) .. (Undo revision 1227906 by Goku20 (talk))

Figura 3.1: Histórico de revisiones de Harry Potter Wiki

Namespace Code	Namespace Name	Notes
{{ns:-2}}	Media	
{{ns:-1}}	Special	Special pages on your wiki
{{ns:0}}	Main	Main namespace for regular wiki articles
{{ns:1}}	Talk	Discussion pages for regular wiki articles
{{ns:2}}	User	
{{ns:3}}	User talk	
{{ns:4}}	Project ("Community Central")	Uses the wiki's sitename
{{ns:5}}	Project talk ("Community Central talk")	
{{ns:6}}	File	
{{ns:7}}	File talk	
{{ns:8}}	MediaWiki	
{{ns:9}}	MediaWiki talk	
{{ns:10}}	Template	
{{ns:11}}	Template talk	
{{ns:12}}	Help	

Figura 3.2: Algunos de los principales namespaces comunes en MediaWiki

a artículos de contenido, vemos un ejemplo en la figura 3.3. Por otro lado, otro namespace que merece la pena que destaquemos es el que engloba las *user talk pages*, estas son páginas propias a cada usuario dentro de la wiki, los usuarios las suelen usar como blogs o como comunicación directa entre ellos, podemos ver un ejemplo en la figura 3.4.

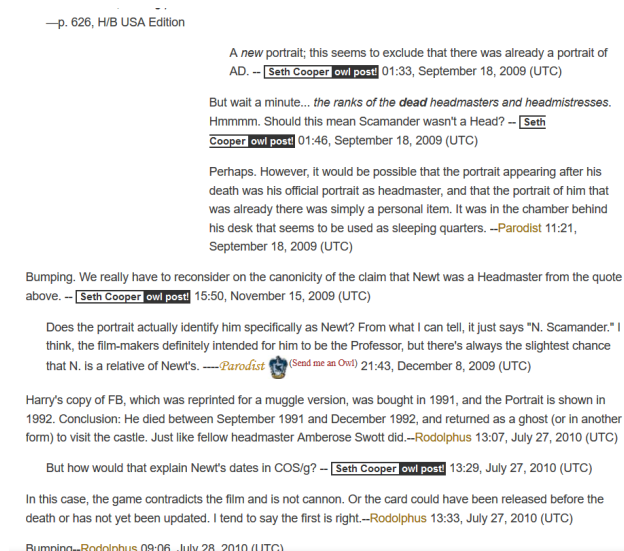


Figura 3.3: Página de talk asociada al artículo; "Newton Scamander" de Harry Potter

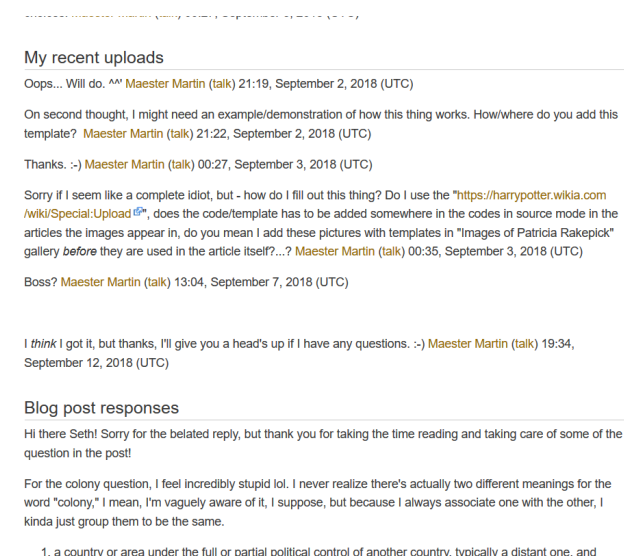


Figura 3.4: Una página de user talk de un usuario en la wiki de Harry Potter

3.2. Discusión de las redes elegidas

Como ya hemos mostrado en la sección anterior 3.1, hay bastantes namespaces que potencialmente podrían ser modelados como una red. La tabla 3.1 sacada de Welser et al. (2011) muestra varios enfoques posibles.

Al final, podemos separar en dos grandes bloques lo que se puede modelar en una wiki, por un lado podemos poner el foco sobre el contenido, o por otro lado, centrarnos en los usuarios (tipo de vértices de la tabla 3.1) y sus relaciones.

Este trabajo pretende estudiar las relaciones entre los usuarios, ver cómo se forman, o qué dinámicas se producen como comunidad. Con lo cual se deja como trabajo futuro (ver 7) el desarrollo de estas otras redes y su estudio (tabla 3.1).

Con el foco puesto ya sobre los usuarios, las relaciones que vamos a querer tener en cuenta van a ser dos; de *Colaboración* y de *Comunicación*. Las primeras como ya hemos hablado en capítulos anteriores van a representar la forma de organización de la comunidad a la hora de producir contenido. Con el segundo tipo vamos a representar, como hace Foote et al. (2019), relaciones más directas entre usuarios, estas atañen a la comunicación que surge entre usuarios para producir contenido, y por otro lado, también vamos a querer reflejar comunicaciones más directas y por tanto más personales entre usuarios.

Otra consideración a tener en cuenta es que vamos a filtrar las aportaciones de usuarios anónimos, estos merecen un estudio a parte que queda fuera de este trabajo. Sabemos que una edición es anónima porque en el histórico queda registrada la IP del ordenador desde el que se realizó la aportación. Cabe aclarar que esta IP puede ser usada por diferentes personas, con lo cual, no está claro que una IP sea equivalente a un usuario.

Por otro lado, también vamos a eliminar todo el contenido que podamos proveniente de bots. Todas las wikis tienen una lista de bots registrados, pero debemos aclarar que no todos los bots están registrados como tal, con lo cual las redes generadas pueden llegar a tener la presencia de algún bot.

Red	Vertices	Aristas	Con peso	Dirigido
Páginas de links	Páginas	Hiperenlaces	Si o No	Si
Páginas de usuario	Usuarios	Comentarios en la página de otro usuario página	Si	Si
Páginas de discusión	Usuarios	Comentario para responder otro comentario de discusión	Si	Si
Página de afiliación de usuarios	Páginas y usuarios (bimodal)	Ediciones de usuario por página	Si	No
Co-Editores	Páginas	Co-Editores	Si	No
Co-Edición	Usuarios	páginas co-editadas	Si	No
Categorías	Categorías	Páginas compartidas	Si	No
Proyectos	Proyectos	Páginas compartidas o usuarios compartidos	Si	No

Tabla 3.1: Tipos de red en base al contenido de la wiki seleccionado.

3.2.1. Definición de la dimensión temporal

Otro aspecto importante que queremos poner sobre la mesa, es acerca de la dimensión temporal, pues las relaciones interpersonales que queremos tratar **no** son fijas en el tiempo. Estas se forman, maduran, y finalmente se diluyen. Tratar de modelar estas relaciones de manera acumulada puede dar una idea de cómo se comporta una comunidad, pero no de cómo ha evolucionado o la tendencia que puede llevar. Más claramente, hablar de comunidades estigméricas u oligárquicas en base a una foto fija, no nos da un conocimiento profundo del porqué sucede o cómo se llega a estas formas de organización.

En nuestro caso vamos a definir la variable temporal de una forma más flexible, es decir, dejamos la libertad al usuario a poder restringir o diluir esta relación entre editores tanto como quiera. Las restricciones que se cumplen son las siguientes:

- Las unidades entre los instantes temporales son de 1 mes. Con lo cual, entre t y $t + 5$ estaremos abarcando un periodo de 5 meses.
- El rango mínimo es entre t y $t + 1$, es decir, un mes. Asumimos de esta manera que en un periodo de 1 mes ha habido colaboración o comunicación entre los editores.

3.3. Definición de la red de comunicación directa entre usuarios

El contenido que define esta red es el generado a partir de las páginas de usuario, estos son espacios en los que los usuarios pueden discutir de una

forma más directa, y nos sirve como canal para ver el flujo de las comunicaciones entre individuos.

Así pues, *la red de comunicación directa* se define tal que; En una ventana V de tiempo, inferimos un enlace desde el editor A hacia el editor B , si A ha editado en la página de usuario de B . De igual manera para inferir un enlace desde B hasta A , B tiene que editar la página de usuario de A . Si A edita su propia página usaremos una métrica propia de esta red como contador de las veces que sucede, esta métrica se puede usar como el tamaño del nodo, lo cual nos dará una idea de la importancia del editor.

De esta forma, *la red de comunicación directa* se define como un grafo dirigido con pesos en las aristas.

3.4. Definición de la red de discusión de artículos como red de colaboración y comunicación

La siguiente red que proponemos se basa en el contenido generado a partir de las redes de discusión, como ya hemos mencionado previamente estas páginas se usan como punto de encuentro entre los usuarios a la hora de editar artículos, a la hora de poner normas de estilo, de estructura o del contenido que esta permitido o no introducir.

Queremos dar un matiz a esta red, ya que la consideramos un nexo de unión entre la red de coedición y la red de comunicación, pues al final lo que se pretende es poner en consenso a los editor sobre un marco común de las reglas que hay que seguir para editar en el artículo X .

Otras representaciones que hemos valorado son; la de Welser et al. (2011) que establece enlaces entre usuarios cuando se replican entre ellos, produciendo así un grafo dirigido. Este enfoque no lo consideramos acertado puesto que lo que se comenta en estas páginas no va solo dirigido a una persona sino a la comunidad como tal.

Otro enfoque es el de Foote et al. (2019), el cual establece un enlace a los 5 usuarios anteriores que hayan editado la página al mensaje del usuario A , esto nos parece un poco arbitrario, puesto que no tiene en cuenta la dilatación temporal que pueda haber entre cada mensaje.

Por lo tanto, definimos *la red de discusión* como sigue; En una ventana V de tiempo, los nodos son usuarios que hayan editado al menos una página de discusión, y enlazamos a dos editores si han participado en la misma página de discusión, si participan en más aumentaremos el peso del enlace en una

unidad. Con esto tenemos un grafo no dirigido.

Como en el resto de redes la ventana V de tiempo es configurable para establecer relaciones más o menos estrechas.

3.5. Definición de la red de Co-edición como red de colaboración

Vamos a definir esta *red de co-edición* en base a los artículos editados, entonces, en una ventana V de tiempo, definimos como nodos a los usuarios que hayan editado al menos un artículo, e inferimos un enlace entre ellos si han editado el mismo artículo. Si ambos usuarios han editado conjuntamente más de un artículo aumentamos el peso del enlace entre ambos en una unidad.

Con lo cual, esta red de colaboración es definida mediante un grafo no dirigido y con pesos.

Dejamos a discreción del usuario el ajustar la ventana temporal V , para definir relaciones de trabajo más estrechas o más laxas, siendo el mínimo intervalo de V de 1 mes. Por ejemplo, en wikis con una actividad muy intensa de ediciones, tener una ventana amplia de tiempo no va a definir correctamente que dos usuarios hayan colaborado en un artículo común, pues es perfectamente posible que el contenido que editó el usuario A haya cambiado completamente cuando el usuario B edita.

Por otro lado, en wikis con una actividad de contenido muy baja, si ajustamos la ventana V al mínimo (un mes), no veremos cambios relevantes, pues hemos aumentado demasiado la granularidad de nuestra red, en consecuencia tendremos redes muy dispersas (los usuarios no se conectan entre sí) o incluso una red vacía en el peor de los casos.

Es remarcable que estos sucesos pueden pasar perfectamente dentro de una misma wiki, pues puede pasar por momentos en los que hay picos de trabajo y momentos de estancamiento.

3.6. Métricas que miden la importancia relativa de los usuarios

Una vez definidas las redes que vamos a emplear para el análisis, se hace necesario disponer de métricas con las que podamos discernir la importancia de los editores en la red. Definimos dos grandes grupos, unas serán métricas propias a la wiki con las que mediremos cuán prolífico es el editor, y otras

irán ligadas a la topología de la red que hemos creado, llamada *centralidad*. El concepto de centralidad va ligado a la facilidad que tiene un nodo (editor en nuestro caso) de ser accedido por otros, una mayor centralidad suele conllevar una mayor influencia en la red de ese nodo.

3.6.1. Métricas de producción

Estas definen un serie de valores sencillos de computar, pero que nos sirven para medir cual ha sido la cantidad de contenido generado por usuario o cuánto ha aportado. Además estas se computan en función de la ventana V establecida previamente.

- **Tenure / Longevidad:** mide la antigüedad del usuario en la comunidad desde su primera edición. Más antiguo implica un valor más alto.
- **Edited Articles / Artículos editados:** es el número de artículos diferentes editados.
- **Article Edits / Ediciones en artículos:** número de ediciones totales en artículos.
- **Edited Talk Pages / Páginas de discusión editadas:** es el número de páginas de discusión editadas.
- **Talk Page Edits / Ediciones en páginas de discusión:** número total de ediciones en páginas de discusión.
- **Edits in its own page / Ediciones en propia página de usuario:** número total de ediciones en su propia página de usuario.
- **User Talks / Páginas de usuario editadas:** es el número de páginas de usuario diferentes editadas.

3.6.2. Centralidad

Como ya hemos dicho las medidas de centralidad miden la importancia relativa de los editores en la red. Nos apoyamos en la aproximación que se dan en Zhang y Wang (2012) para darlas contexto.

- **Degree / Grado:** nos da el número de enlaces que inciden sobre un editor. Si el grafo es dirigido tendremos **In-degree** como grado de entrada, y **Out-degree** como grado de salida del editor. Con esta métrica podemos medir el número de colaboraciones que involucran al editor.

- **Closeness / Cercanía:** mide la facilidad de acceso al resto de la red, o dicho de otro modo, se busca que la distancia al resto de nodos de la red sea pequeña. Lo asociamos con un rol generalista, un valor elevado indicará que ese editor diversifica su trabajo en la comunidad en vez de centrarse en solo una parte. Toma valores en el rango $(0, 1]$
- **Betweenness / Intermediación:** mide la centralidad desde el punto de vista de cuántos caminos mínimos pasan por el editor. Lo asociamos con el rol de colaborador-puente, pues cuando tiene un valor muy alto significa que trabaja con editores de diferentes partes de la wiki. Toma valores en el rango $[0, +\infty]$ para grafos con pesos.
- **Page Rank:** se basa en el algoritmo de búsqueda de Google, y mide no solo la cercanía de un editor a otro sino también la de los vecinos más próximos. Así un editor que pueda parecer aislado en un principio, si esta conectado a otro nodo con una betweenness alta este le dará acceso más fácil al resto de la wiki. Toma valores en el rango $[0, 1]$

3.7. Métricas que muestran el comportamiento de la comunidad

Para dar un sentido más general a las redes también disponemos de métricas en las que apoyarnos y que nos sirven muchas veces de resumen de las métricas locales a los nodos.

- **Density / Densidad:** mide cuánto de conectada esta la red, un valor alto demuestra una cohesión alta entre los editores de la red, es decir, donde todos los editores trabajan entre sí. Toma valores en el rango $[0, 1]$
- **Gini of Degree / Gini de grado:** el coeficiente de Gini mide la desigualdad en la red, aplicándolo al grado de los editores nos da un valor con el que podemos medir la desigualdad en las colaboraciones. Si nos da un valor cercano a 0 podemos decir que es una comunidad en la que la participación es equitativa entre todos los artículos. Por el contrario, un valor cercano a 1 indicaría que un grupo pequeño de editores son los que centralizan las comunicaciones o colaboraciones. Toma valores en el rango $[0, 1]$
- **Gini of Betweenness / Gini de intermediación:** nos sirve para medir la desigualdad del rol de colaborador-puente, un valor cercano a 0 indicaría que este rol no es apreciable en la comunidad. Pero si topamos con un valor cercano a 1, querrá decir que este rol es la llave en la red. Toma valores en el rango $[0, 1]$

- **Assortativity Degree / Asortatividad de grado:** como se plantea en Barabási y Pósfai (2016), es el coeficiente de correlación que mide como se asocian los editores entre sí. Toma valores en el rango $[-1, 1]$. Si se encuentra en el rango de:

$$\mu = 0,37 \pm 0,11$$

decimos que la red es *asortativa*, es decir, que los editores que son hubs (tienen un grado alto) tienden a juntarse entre ellos, y los que tienen un grado más bajo tienden a juntarse entre editores con el mismo grado.

Por otro lado, si el coeficiente se encuentra en el rango de:

$$\mu = -0,76 \pm 0,04$$

decimos que la red es *disortativa*, es decir, los hubs tienden a agruparse con los nodos con menor grado y se repelen entre ellos.

Por último, con un valor cercano a 0, se dice que la red es *neutra*, y los enlaces se producen de forma aleatoria.

Proporcionamos además otras métricas adicionales que pueden ser de ayuda en el análisis, como el diámetro de la red con el que podemos saber la distancia máxima entre nodos, el número de componentes conexas, o las páginas editadas y el número total de ediciones en función de la red que se disponga.

3.7.1. Búsqueda de comunidades

A menudo en las redes complejas se tienden a producir subcomunidades debido a la disposición de los enlaces, los cuales forman zonas altamente conectadas en la red y otras menos densas. Una comunidad ideal es la que forma un clique (componente conexa en la que todos los nodos se conectan entre sí), si varios de estos cliques se conectan eventualmente por algún enlace, entonces es fácil de ver que cada clique forma su propia comunidad a pesar de ser todos alcanzables entre sí.

El problema de buscar comunidades en base a cliques es que es un problema NP-Completo, y a parte, es también una forma más rígida de establecer comunidades (todos con todos), con lo cual si queremos tener un algoritmo que no sea tan costoso ni tan riguroso al establecer comunidades, entonces debemos relajar la condición de comunidad y tendremos que buscar alternativas.

En nuestro caso nos hemos basado en el algoritmo que se describe en Pons y Latapy (2005). La idea bajo el algoritmo es que caminos aleatorios cortos tienden a mantenerse dentro de una misma comunidad. Así, por cada nodo se lanzan varios caminos aleatorios y finalmente se hace una ponderación de con que vecinos se ha mantenido más cercanía durante la simulación.

A la hora de estudiar las comunidades que forman las wikis podemos usar este algoritmo ver si se forman pocas subcomunidades o clústeres, lo que indica una colaboración más próxima entre todos los usuarios, o por el contrario muchas comunidades significa que tenemos grupos de trabajo separados.

Capítulo 4

Arquitectura de la aplicación y modelo de proceso

En el presente capítulo hablaremos de cómo hemos construido *WikiChron networks*, tecnologías sobre las que se sustenta, integración en el ecosistema *WikiChron*, el proceso de extracción, transformación y cargado (load) de los datos (ETL). Terminaremos mostrando las ventajas y limitaciones técnicas que presenta.

El repositorio se encuentra en: <https://github.com/Grasia/WikiChron>, y lo hemos desplegado en producción para una mayor comodidad de uso a la hora de probarlo, <http://wikichron.science>.

Conviene que vayamos adelantando que no todo el código incluido en el repositorio (el mencionado más arriba) es parte de este proyecto, más concretamente, lo que sí se ha desarrollado para el proyecto es todo lo que cuelga de: **WikiChron/wikichron/dash/apps/networks/**. Se hará una referencia explícita cuando algo no haya sido desarrollado para este trabajo.

4.1. Tecnologías

Desde un inicio teníamos en mente que el desarrollo de la herramienta tenía que venir de la mano con una integración en *WikiChron classic*, herramienta desarrollada por Abel Serrano y Javier Arollo para el grupo de investigación *GRASIA* de la *Universidad Complutense de Madrid*, que trata el mismo problema pero desde un punto de vista de analíticas y series temporales.

Por lo tanto, para hacer esta integración efectiva desde un primer momen-

to, se adoptaron las mismas tecnologías usadas para el desarrollo de *Classic*, esto es, el uso de *Python 3.6* sobre el framework *Plotly-Dash*. La ventaja de usar este framework es que nos permite programar tanto el backend como el frontend sobre un único lenguaje, *Python*.

Los datos que se manejan en el programa son en última instancia csv's o ficheros de texto plano, los cuales son manipulados mediante la librería de *Pandas*. Estos ficheros contienen un histórico de las ediciones de cada wiki con valores separados por ';'. Aunque este recorrido ya estaba allanado previamente por Abel en el desarrollo de *WikiChron-classic*, y por lo tanto, nos basaremos en este principio para construir las redes discutidas en el capítulo anterior.

A partir de aquí, tuvimos que elegir diferentes tecnologías para el manejo y visualización de los grafos. En primer lugar, para manipular los datos en forma de grafos y calcular sus métricas asociadas discutimos diferentes alternativas, aunque eso sí, teniendo siempre en mente usar una tecnología extendida y bien documentada. Entre las opciones destacamos *NetworkX* y *python-igraph*. Finalmente se optó por el uso de *python-igraph* por dos motivos: una es que esta librería está hecha en C, con lo cual cálculos o algoritmos más costosos son más llevaderos que en *NetworkX* desarrollado íntegramente en Python.

Otra razón para inclinarnos por *python-igraph*, era que ya habíamos usado *NetworkX* anteriormente, y buscábamos aprender nuevas tecnologías.

Por último, para la presentación del grafo en la app tuvimos muy pocas opciones, por un lado estaba; *D3.js*, librería muy potente para gráficos pero no especializada en grafos y con la necesidad de traducir código JS a código Python. Otra opción a tener en cuenta era el uso de *Plotly*, aunque sí que era usable directamente con Python no se especializa en el uso de grafos. Ambas opciones suponían un incremento sustancial en el tiempo de desarrollo, lo cual, no era admisible si queríamos llevar a la práctica todas las ideas que teníamos en mente.

Finalmente, dimos con *dash-cytoscape* una implementación de la librería *Cytoscape.js* sobre el framework *Plotly/Dash*, ahora sí pudimos disponer de una librería especializada en el manejo de grafos. Aunque esta se encontraba y encuentra en un estado inicial de su desarrollo, por tanto, aún presente algunos fallos menores que han sido solucionados o lo serán en un futuro.

4.2. Arquitectura

En la figura 4.1 queda reflejada de forma esquemática la arquitectura de las tecnologías usadas. Y en la figura 4.2 la correspondencia con los principales paquetes de código involucrados.

En líneas generales hemos respetado la arquitectura previa montada para *WikiChron-classic*, esto es, estructura de paquetes y la comunicación entre ellos. Aunque dado que los requisitos son cambiantes y hay que introducir siempre nuevos archivos y paquetes esto obviamente ha cambiado hasta lo mostrado en 4.2.

Vamos a dar una visión general de como funciona la aplicación. En primer lugar, se llama a la aplicación con los códigos de la wiki, red, e intervalo temporal que se quiere mostrar (se hace una selección previa de estos valores, o porque se ha compartido un enlace). Con esto se llama al archivo **data_controller.py** para que devuelva el dataframe asociado a la wiki. Este además se filtra temporalmente y para limpiar bots y usuarios anónimos. A continuación, se elige una red del paquete **networks/models**, se crea la red correspondiente en base al dataframe y se calculan las métricas asociadas.

El proceso anterior termina con un grafo de la librería *igraph-python*, pero para mostrarlo usamos la librería *Cytoscape*, con lo cual, hay que parsear el grafo a un formato que se pueda entender, más concretamente, a un diccionario.

Finalmente, tenemos solo falta componer la hoja de estilos del grafo a mostrar, para ello, usamos el archivo **CytoscapeStylesheet.py** que nos sirve como fachada de las operaciones que queremos permitir para hacer cambios al estilo. Con esto ya se actualiza la pantalla y los datos de las métricas calculadas.

Cabe resaltar que cada vez que se cambia el tipo de red, la wiki o la ventana temporal se vuelve a ejecutar todo lo anterior. No pasa lo mismo si deseamos hacer cambios de estilo sobre la red, como cambiar de tamaño los nodos o color (esto se detalla en el siguiente capítulo), pues solo se recalcula la hoja de estilos.

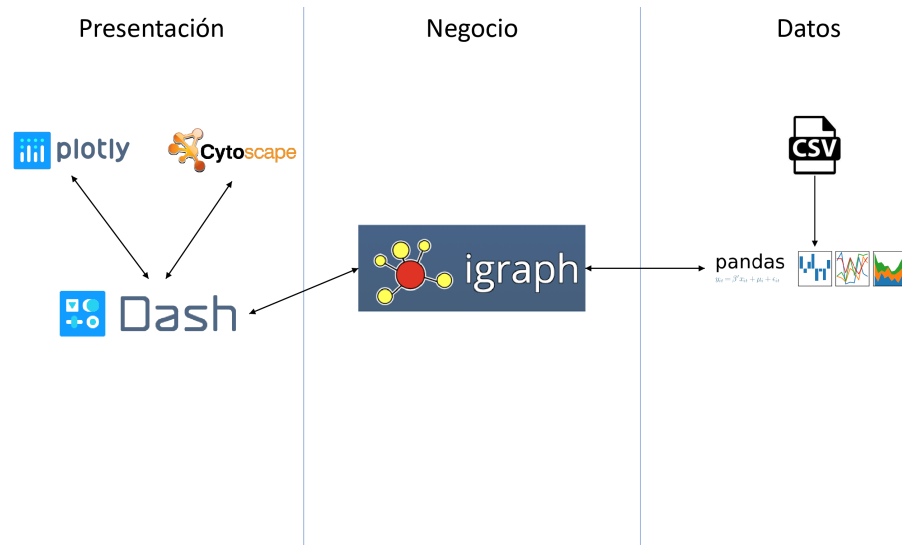


Figura 4.1: Modelo de tres capas de las tecnologías usadas

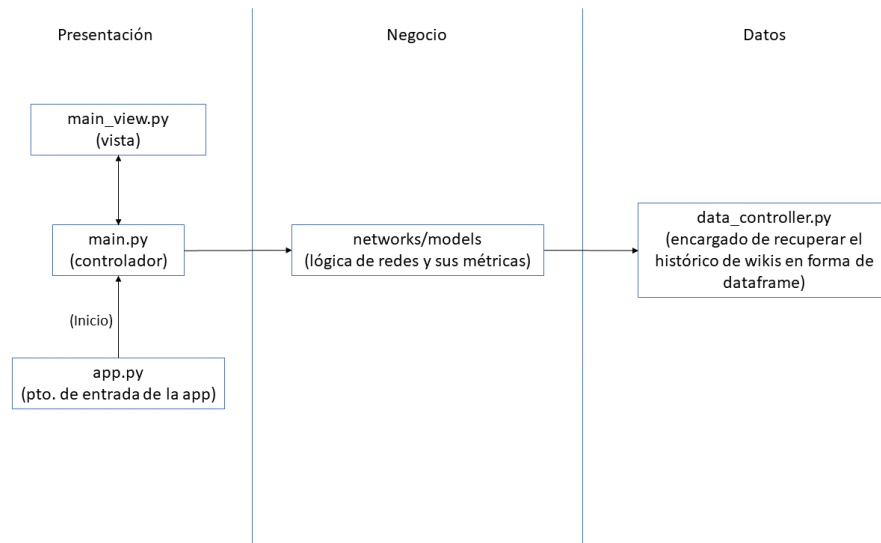


Figura 4.2: Modelo de tres capas de los archivos y paquetes involucrados. Estos se encuentran en la ruta **WikiChron/wikichron/dash/apps/networks**

4.2.1. Proceso ETL de los datos

Para la extracción de los datos, las wikis que se utilizan sobre el software de *MediaWiki*, guardan un histórico de los datos en formato xml¹.

Para la descarga y manejo de estos datos vamos a seguir el proceso que

¹Más información en: <https://www.mediawiki.org>

ya estaba implementado por Abel Serrano en *WikiChron classic*.

En primer lugar, para la extracción de los datos debemos llamar a la API, el principal problema de esto, es que la api nos limita a un máximo de 5,000 ediciones por llamada, con lo cual hay que descargar el contenido de la wiki en partes que luego hay que unir. Esto lo hemos realizado con los siguientes [scripts](#).

Una vez descargado el histórico, tendremos un archivo xml bastante pesado, lo cual es impracticable de manejar. Por lo tanto, tenemos que procesar el archivo para reducir su tamaño y transformarlo a un formato más manejable, en nuestro caso csv. Terminado el proceso pasaremos de un archivo xml de gigas a un archivo csv de megas.

En la aplicación, los ficheros csv resultantes se tienen almacenar en la carpeta **WikiChron/data**/² de nuestro proyecto, para servirlos a la aplicación. El proceso se realiza con la siguiente librería [wiki-dump-parser](#) (desarrollada por Abel Serrano).

Por último, se precisa de un archivo "**wikis.json**" en la ruta **WikiChron/data/** que almacene metadatos de todos los csv de wikis disponibles, como puedan ser los id de bots, enlaces a las wikis o última fecha de la actualización del archivo. Para generar este archivo hacemos uso del script **WikiChron/scripts/generate_wikis_json.py**.

4.3. Modelo de proceso

Todo desarrollo tiene un modelo de proceso, y este proyecto no es en absoluto una excepción. Hemos querido dedicar una sección al proceso que ha dado vida a *WikiChron networks*, debido a que finalmente ha involucrado a diferentes personas, que aún no siendo beneficiarios últimos, si que se han comprometido con el proyecto y han ayudado mucho a sacarlo adelante.

4.3.1. Primera fase

En un principio, queríamos trabajar sobre el fenómeno de las wikis pero no concebíamos el enfoque concreto. Se comenzó por acudir al *TFM* de Villa (2018) para ver de primera mano si podía servirnos como punto de partida. Además, dado que el alumno firmante de este *TFG* se encontraba matriculado en la asignatura de "*Análisis de redes sociales*"(SOC), se decidió más tarde que era una buena idea presentar como trabajo final de la asignatura

²Hablamos siempre del repositorio del proyecto: <https://github.com/Grasia/WikiChron>

lo que podría ser una versión temprana del sistema.

Para esta primera fase, puesto que el tiempo era limitado, propusimos una especificación de requisitos mínima del sistema, esto era, tener una red para trabajar y poder calcular algunas métricas sobre ella.

En cuanto al *diseño*, desde un inicio se quiso integrar el nuevo sistema, con el ya desarrollado *WikiChron classic*, con lo cual, se decidieron las tecnologías usadas como ya hemos discutido en 4.1. Además en colaboración con Guillermo Jiménez, Javier Arroyo y Abel Serrano, desarrollamos un modelo de la red a incluir en el sistema (finalmente descartado).

Finalmente, las fases de *implementación* y *validación* fueron concurrendo a lo largo del siguiente mes, desembocando en última instancia en lo que se muestra en la figura 4.3.

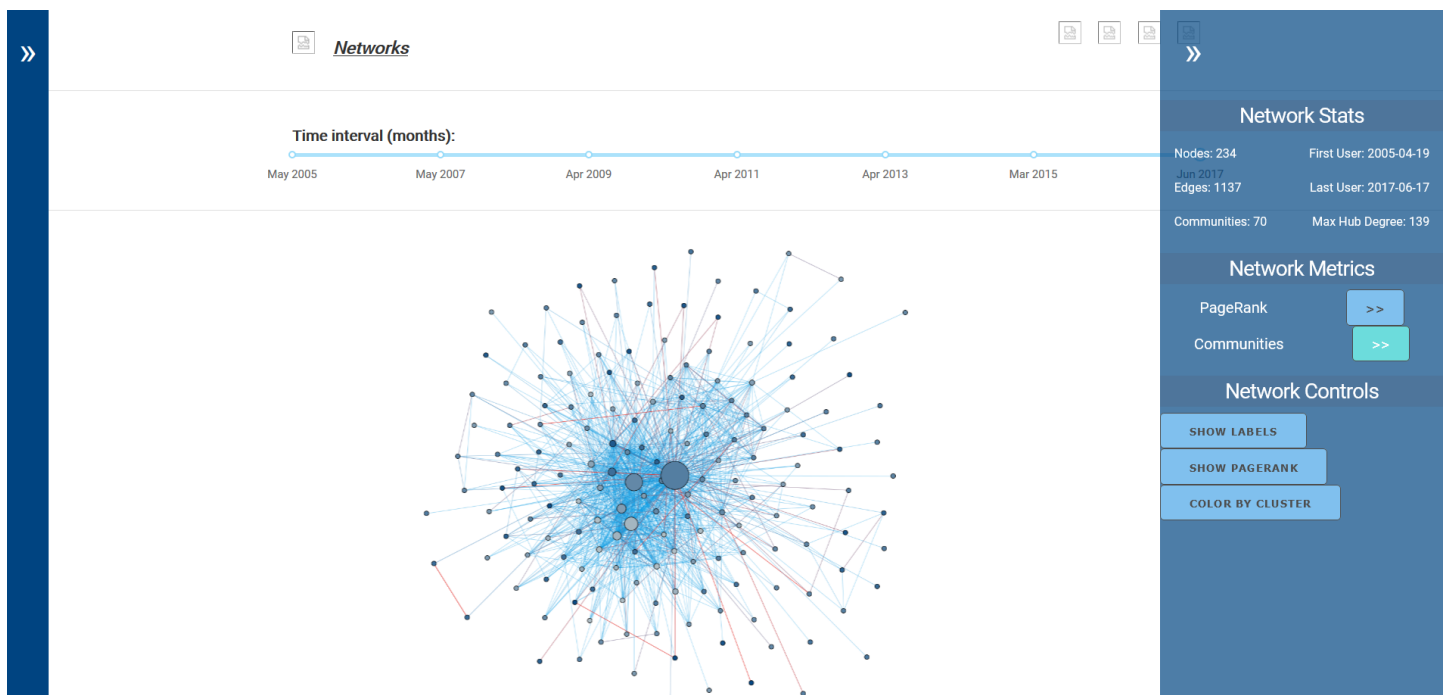


Figura 4.3: Primera versión de WikiChron networks

4.3.2. Segunda fase

Superada la primera fase, que vino dada por una metodología en cascada, se decidió que teníamos que optar por un desarrollo más ágil con todo lo que

implicaba.

Se comenzó por una fase de refactorización del código, con esto teníamos como objetivo pasar de una implementación monolítica a una implementación que fuera algo más flexible, teniendo en mente siempre que íbamos a tener una especificación de requisitos incremental y cambiante. Para ello se implementaron diferentes patrones de diseño que facilitaran la reusabilidad y el mantenimiento.

A partir de aquí, comenzamos a establecer tareas semanales o quincenales, así como reuniones semanales con Abel Serrano para poder integrarlo todo en un mismo ecosistema, pues manteníamos un desarrollo paralelo de ambas herramientas y era vital tenerlo todo sincronizado. Se establecieron reuniones del mismo modo, para definir y discutir los nuevos modelos de red que describieran mejor el comportamiento de las comunidades.

Además, para el papel de tester/cliente tuvimos eventualmente la ayuda de Guillermo Jiménez, que nos ayudó a definir los requisitos necesarios del sistema y probarlo.

También dispusimos de la ayuda de Elena Martínez Vicente, que diseñó la interfaz de la aplicación principal para integrarla con coherencia con *WikiChron-classic*.

De esta forma, realizamos una integración continua para que las pruebas detectasen tanto bugs como problemas en la interfaz o cambios en los requisitos.

Finalmente, este proceso no permitió ser más efectivos a la hora del desarrollo, y poder escribir un artículo para *OpenSym 19* contando el desempeño de la herramienta y dando casos de uso prácticos, El Faqir et al. (2019).

4.4. Ventajas y limitaciones técnicas

En cuanto a ventajas, tenemos que la herramienta se integra bastante bien en el entorno *WikiChron*, además el diseño final permite disponer de toda la información de un vistazo. Por otro lado, un "huevo de Colón" ha resultado ser el poder compartir enlaces con otros, por ejemplo, a la hora de escribir el *paper* El Faqir et al. (2019) nos fue de gran ayuda el poder compartir los resultados para ejemplificar nuestras argumentaciones.

Otra buena aportación es el poder descargar los datos de la red que se están visualizando, de esta manera si se quiere usar otra herramienta como *Gephi* o *Cytoscape* solo hay que importar el archivo GML resultante, el cual, es un formato orientado a grafos, ampliamente soportado por las diferentes herramientas y librerías disponibles en el mercado.

La gran limitación técnica que tiene la herramienta, es que bajo carga de grandes cantidades de datos el rendimiento se resiente. Esto puede suceder por las dimensiones de la wiki elegida, o porque hayamos elegido rangos temporales demasiado amplios con lo cual, la fluidez para generar el grafo se resiente. La mejor forma de combatir esta limitación es ver de ante mano en la pantalla de selección las estadísticas de la wiki, y si esta es muy grande, entonces tendremos que establecer rango temporales más limitados. Además como ya discutimos en el capítulo 3, ampliar la ventana temporal reduce también el entendimiento de la red, puesto que las relaciones que se establecen son demasiado laxas.

Capítulo 5

Guía de uso

Si entramos en <http://wikichron.science/>, encontraremos algo parecido a lo que se puede ver en la figura 5.1, donde se pueden elegir los diferentes modos de *WikiChron*.

Si seleccionamos *WikiChron networks*, pasaremos a la siguiente pantalla (figura 5.2), donde podemos elegir la wiki que queremos visualizar, la red sobre la que queremos trabajar, y las fechas sobre las que filtrar. Siendo el único requisito para progresar elegir la wiki (recalcamos que las pantallas 5.1 como 5.2 no pertenecen al ámbito de este proyecto).



Figura 5.1: Página de inicio de WikiChron

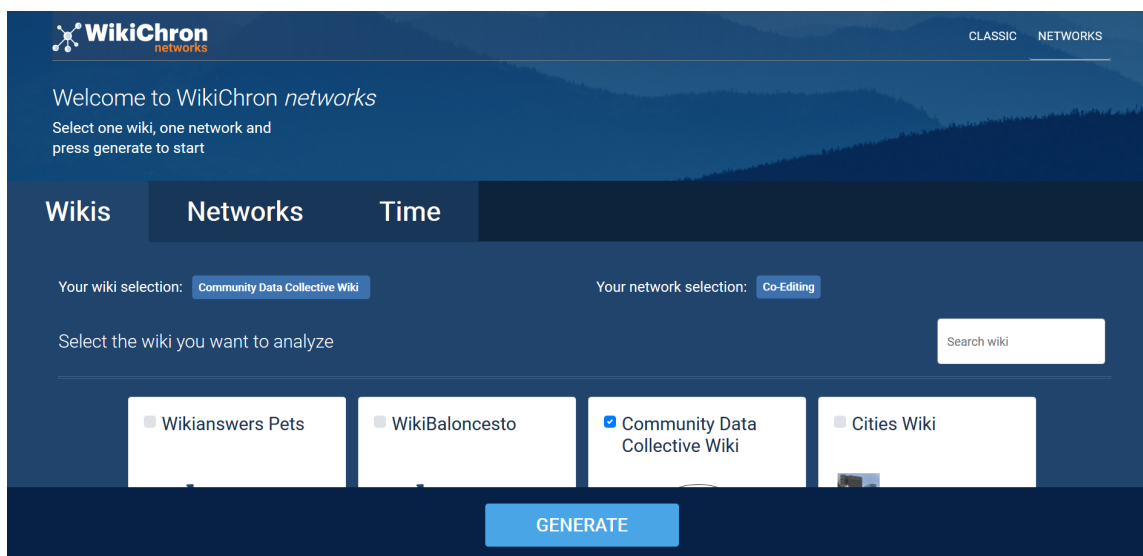


Figura 5.2: Página de selección previa

Una vez hecha la selección, si pulsamos sobre el botón de "Generate", tras unos instantes¹ se generará la red como se muestra en la en la figura 5.3.

Si empezamos por la cabecera, tenemos un par de enlaces (en la parte izquierda) para volver atrás en la selección de la wiki, y para cambiar rápidamente entre entre las diferentes redes para la misma wiki seleccionada. Por el lado derecho, tenemos dos botones, uno con el que podemos compartir el enlace a esta wiki con la misma selección temporal. Y el otro botón que nos sirve para descargar los datos ya calculados de la red en formato *GML*.

En la parte superior, del cuerpo de la página (donde sale el título de la wiki), disponemos de la selección temporal, podemos jugar con los valores para filtrar temporalmente la red. Los "steps" del slider son mensuales como indican las etiquetas que se encuentran encima de las asas. Además podemos manejar el slider con los controles incluidos inmediatamente debajo, los cuales nos permiten manejar ambas asas en base al número de meses que las queramos mover a un lado u al otro.

Si seguimos el orden descendente, debajo del filtro temporal se disponen los diferentes controles de la red. Podemos mostrar u ocultar la leyenda, las etiquetas de los nodos, o resetear la vista del grafo. Además, otros controles importantes son el de mostrar las comunidades, o el de colorear y dar tamaño a los nodos en base a una métrica (respectivamente en 5.4 y 5.5).

¹El tiempo de generación puede variar mucho entre wikis, y también por el rango temporal seleccionado

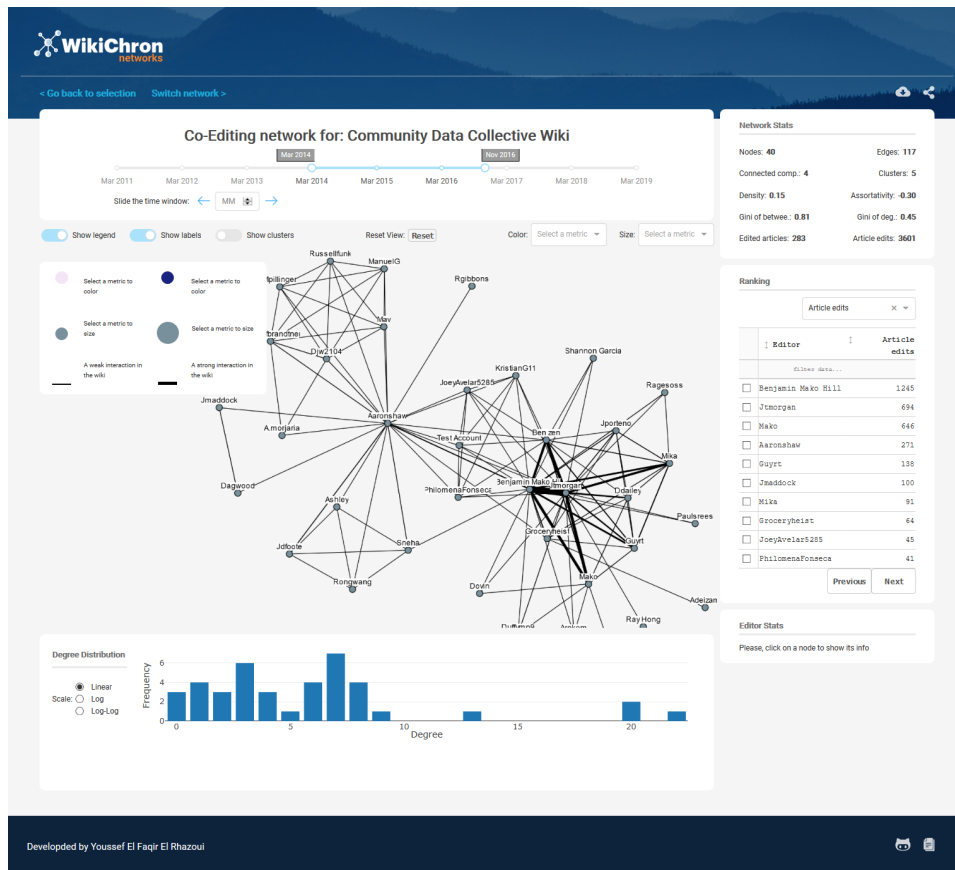


Figura 5.3: Red de co-edición de la wiki Community Data Collective Wiki

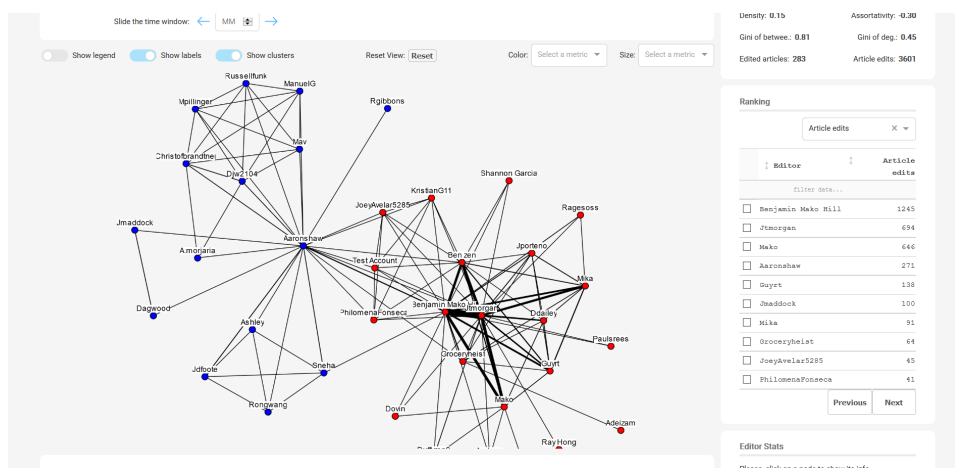


Figura 5.4: Clusters coloreados de diferentes colores

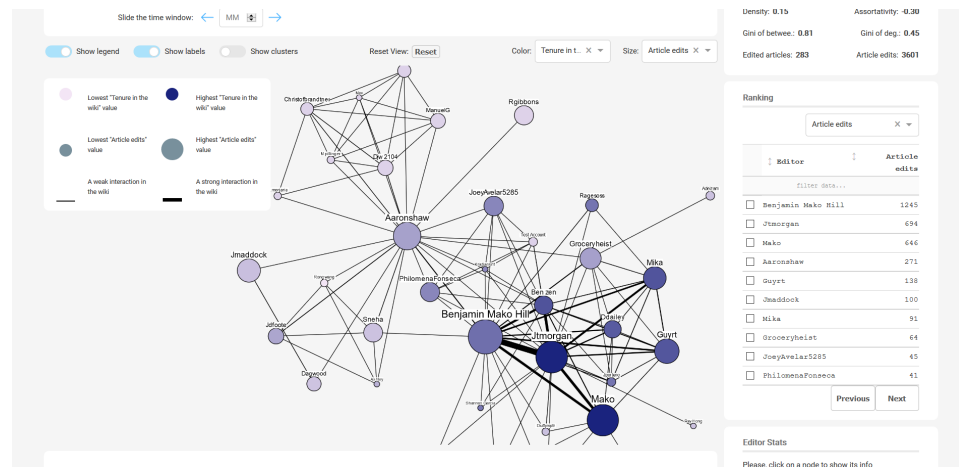


Figura 5.5: Degradado de la longevidad de los usuarios. Azul oscuro indica usuarios más antiguos. El tamaño refiere a la cantidad de ediciones en artículos.

En el centro de la página nos encontramos con la red, esta se presenta mediante un layout de fuerzas, estos tipos de layout simulan una interacción mediante fuerzas en función del tamaño del nodo y las aristas, éstos por lo general dejan en el centro a los nodos más grandes y escoran a los nodos menos relevantes a la periferia. Por otro lado, podemos interactuar con la red, haciendo zoom y moviendo nodos, si pinchamos sobre una arista se nos muestra encima de esta el peso que tiene. Si queremos restablecer la vista podemos pulsar sobre el botón de "reset" mencionado anteriormente.

En la parte baja de la página disponemos de un gráfico de barras de la distribución de grados de los nodos, podemos cambiar a diferentes escalas. Es útil para ver el comportamiento de la red, por ejemplo si sigue una ley potencial o una distribución de *Poisson*.

En la parte superior de la columna lateral de la página, se presenta un panel con las métricas globales ya descritas en el capítulo 3.7.

El siguiente panel muestra una tabla con el top diez de usuarios en base a una métrica seleccionada. La tabla se puede paginar, ordenar por columnas, y filtrar. Para filtrar nombres pondremos **eq "Nombre"**, y para filtrar valores pondremos por ejemplo: **>"50"** para mostrar solo los valores mayores a 50.

Por último, si seleccionamos una fila en la tabla se nos marcará el nodo asociado en la red. Además si pinchamos sobre cualquier nodo de la red aparecerá su información en el último panel de la barra lateral (figura 5.6).

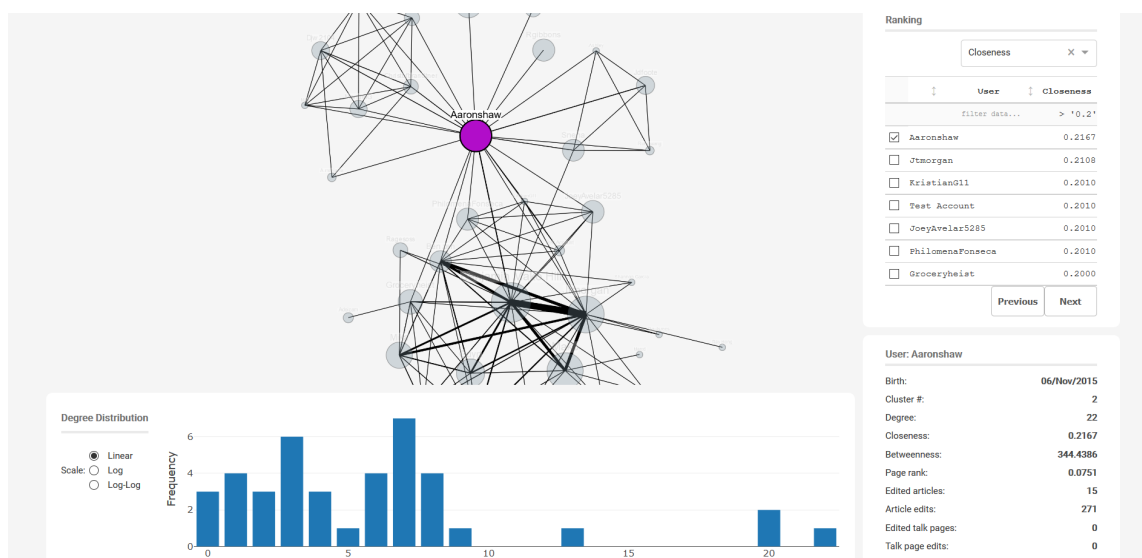


Figura 5.6: Selección de una fila de la tabla, e información concreta de un nodo.

Capítulo 6

Análisis

En el siguiente capítulo proponemos un caso de estudio, con el fin de mostrar la utilidad de la herramienta. Además añadiremos links que permitan comprobar de primera mano las afirmaciones propuestas, pues el lector podrá acceder mediante el link a la aplicación.

6.1. El caso de Cities Wiki

Esta es una wiki de habla inglesa en la que se recoge la historia y cultura de diferentes ciudades del mundo. Hemos querido seleccionar dicha wiki para poder observar si se producen sub-comunidades, en concreto, queremos analizar si existen distintos grupos de editores que trabajan en distintas partes de la wiki, lo cual podría decir, que gente de diferentes culturas/regiones trabajan de manera aislada.

Esta es una wiki con unos 15 años de historia, con lo cual vamos a poder ver la evolución que ha tomado, teniendo así periodos con mucha actividad y otros con menos actividad.

6.1.1. Primera etapa

Empezamos pues con una visión global de la primera etapa, que hemos centrado entre [Nov 2006 - Nov 2007](#), en la figura 6.1 se muestra parte de la red de coedición resultante, donde el tamaño de los nodos corresponde con el número de ediciones realizadas en artículos, y el color refleja la pertenencia a una comunidad. Es visible la formación de diferentes comunidades, y que el usuario *Robin Patterson* parece el coordinador principal, pues comunica con las diferentes comunidades formadas y partes de la red. Además el tamaño nos indica cuales son los editores más entusiastas. Aún con todo considera-

mos que la ventana temporal tomada es muy laxa, 1 año es demasiado para considerar que dos individuos hayan podido colaborar.

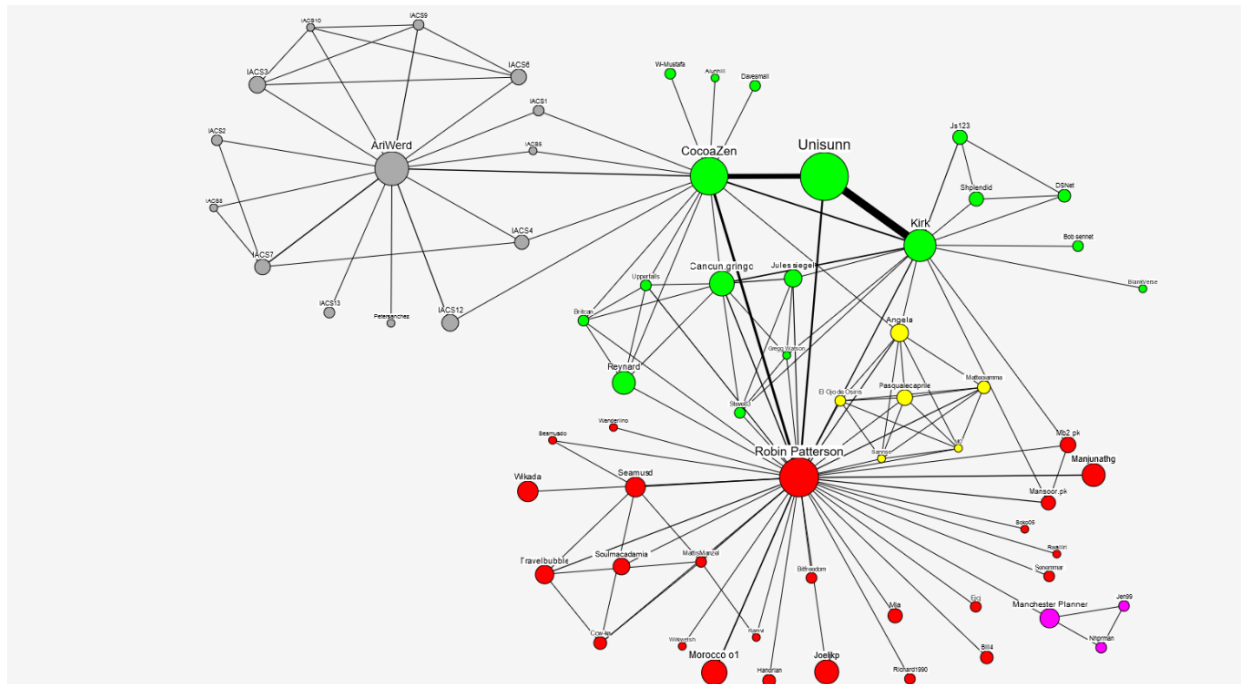


Figura 6.1: Red de co-edición de artículos entre Nov 2006 - Nov 2007

Pasamos entonces a aumentar la lupa de nuestra red al periodo de [Nov 2006 - Feb 2007](#), que comprende un rango de 4 meses entre ediciones. En la figura 6.2 podemos ver una red un tanto dispersa pues para **39 usuarios** apenas tenemos una densidad del **0.04**, lo que implica que haya muy poca cohesión como grupo, pues cerca del **40 %** de los editores emprenden proyectos que no involucran a más gente.

Podemos notar que tanto el usuario *Robin Patterson* como *CocoaZen* son los encargados de gestionar y coordinar la wiki, esto es confirmable con el hecho de que el gini de betweenness es de **0.94**, un valor muy alto que nos indica la desigualdad tan elevada que hay a la hora de conectar las diferentes partes de la red, lo que llamamos en su momento "colaborador-puente", dicho de otra manera, muy poca gente es la encargada de conectar al resto de usuarios de la red entre sí. Además, el valor de assortativity degree (**-0.43**) nos indica una ligera tendencia hacia las disasortatividad, es decir, los hubs tienden a evitarse entre ellos y los editores más casuales (con poco grado), se conectan a dichos hubs. Todo ello refuerza la idea de ser una comunidad sostenida sobre dos usuarios principales.

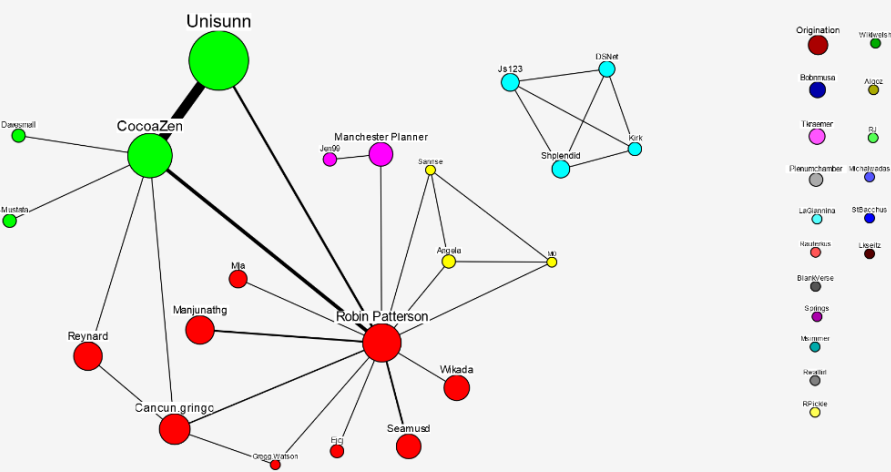


Figura 6.2: Red de co-edición de artículos entre Nov 2006 - Feb 2007

En la tabla 6.1 podemos ver un top 5 de los usuarios con las métricas más altas de la red. Destacamos el papel de *Unisunn* que teniendo unos valores tan bajos en métricas que miden su importancia relativa en la red, tiene unas métricas de creación de contenido altísimas, esto da a entender que es un usuario que se ha involucrado en un proyecto propio dentro de la wiki, y apenas se comunica con los usuarios centrales de la misma.

Usuario	Degree	Closeness	Betweenness	Artículos editados	Ediciones totales
Robin Patterson	13	0.0444	117	44	52
CocoaZen	6	0.0432	31	76	105
Cancun.gringo	4	0.0437	52	7	21
Unisunn	2	0.0413	0	276	594
Manjunathg	1	0.0421	16	3	16

Tabla 6.1: Un top de usuarios con las métricas más altas de la red de co-edición de artículos entre Nov 2006 - Feb 2007

En cuanto a la red de [discusión](#) en el mismo periodo, vemos una red muy pobre tan solo los coordinadores principales se comunican a través de ella. Esto nos indica que solo los usuarios más veteranos o coordinadores principales discuten acerca de como producir contenido, lo que por otro lado puede

tener sentido pues no es una wiki con muchos artículos y una necesidad clara de coordinación.

Si pasamos a la red de [comunicación directa](#) del mismo periodo, podemos ver en la figura 6.3 (el tamaño de los nodos refleja ediciones en su propia página de usuario), la presencia de principalmente 2 clústeres en forma de estrella, sustentados sobre los usuarios *Robin Patterson* y *CocoaZen*, esto se debe a una política de recibimiento de nuevos usuarios que se mantenía en esa época para captar y mantener a nuevos miembros ([un mensaje de CocoaZen a Manchester_Planner](#)). Podemos confirmar además, el hecho de que la mayoría de comunicaciones son de unos pocos usuarios a la mayoría viendo que el gini out-degree es de **0.97** y que el gini betweenness es de **0.98**.

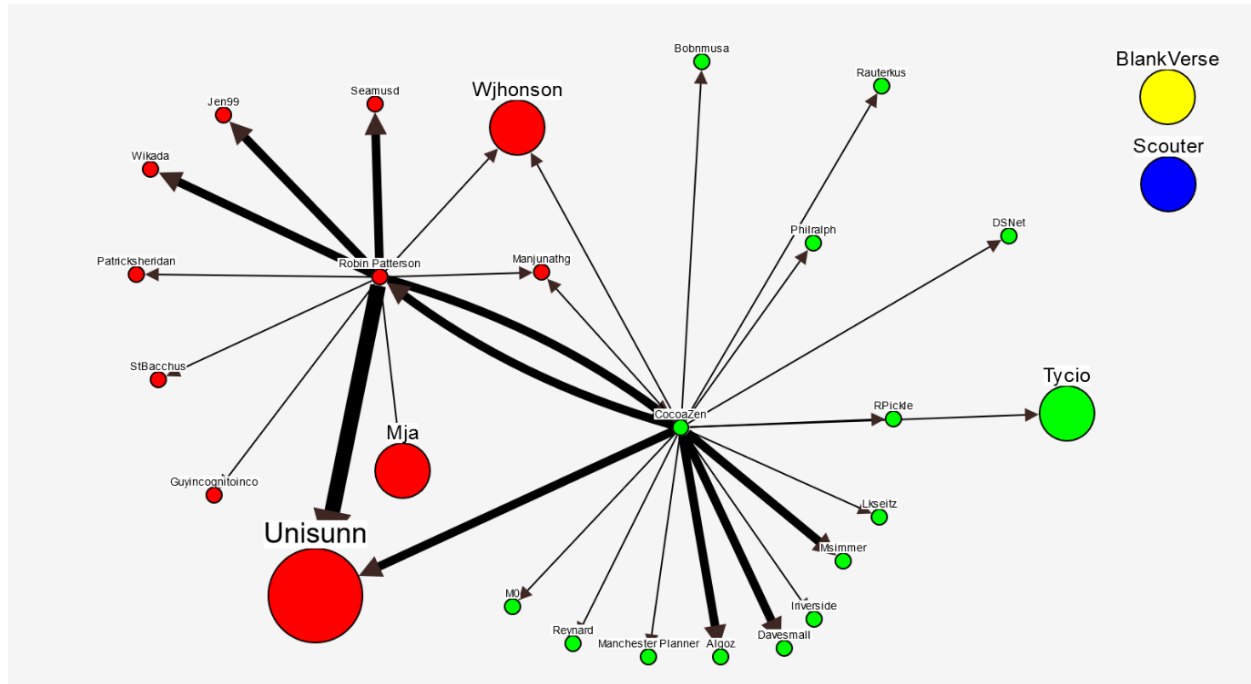


Figura 6.3: Red de comunicación directa entre Nov 2006 - Feb 2007

6.1.2. Segunda etapa

Esta etapa va a venir dada por un estancamiento tanto a la hora de producir contenido nuevo, como a la hora de atraer usuarios nuevos. Ahora nos centaremos en la red de coedición en el periodo que comprende [Nov 2010 -](#)

Feb 2011.

En la figura 6.4 podemos ver una red con una estructura más pobre, apenas se han realizado **33** ediciones en 4 meses, además el número de usuarios activos a descendido hasta los **13** que conforman la red. Es destacable que los dos usuarios que antes se encargaban de la coordinación de la wiki en este periodo han pasado a un papel muy secundario.

Por otro lado, se ha formado un cluster de usuarios que editan páginas de ciudades pakistaníes, estos conforman en su mayoría el contenido generado durante estos meses. En su conjunto la red conforma una densidad del **0.21**, y el coeficiente de gini de betweenness nos da **0.75**, seguramente estos datos fueran más equitativos si contáramos solo con el grupo de editores paquistaníes, pues parece que se coordinan bastante para editar.

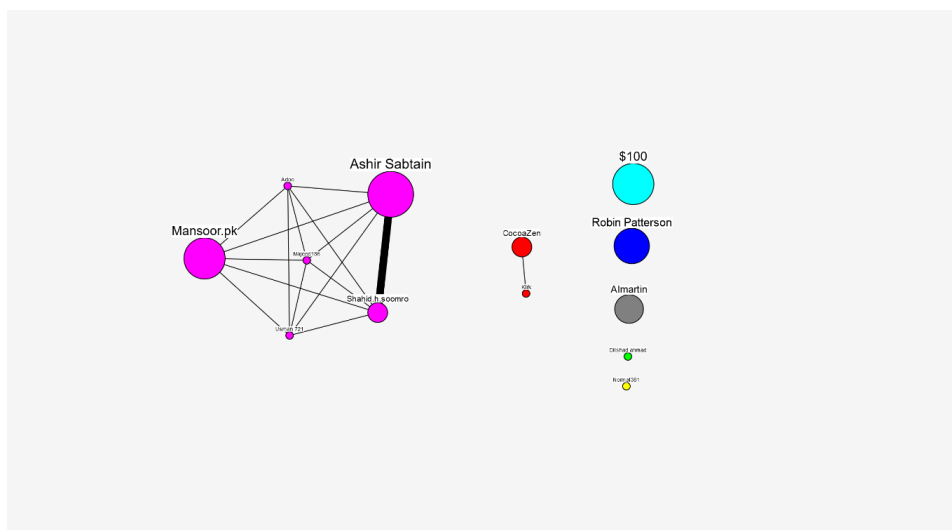


Figura 6.4: Red de coedición entre Nov 2010 - Feb 2011

En cuanto a las redes de discusión y de comunicación directa no se generan ediciones en ese periodo. Aunque si que es verdad que en esta wiki las páginas de discusión no han sido muy usadas en general, incluso en la etapa anterior que era más boyante. Además donde en la etapa anterior había una fuerte presencia de los usuarios coordinadores que mantenían una política de integración de los nuevos usuarios, ahora se han abandonado por completo estas políticas, lo que repercute en una menor entrada de usuarios nuevos y finalmente en menor contenido generado en beneficio de la wiki.

Ya por último, en el periodo que comprende [Nov 2016 - Feb 2017](#) vemos que la wiki está prácticamente desierta, apenas hay **6 editores** para **14 edi-**

ciones que hayan participado en los 4 meses seleccionados, en la figura 6.5 vemos una relación de trabajo casi nula. Además, parece que tanto *Robin Patterson* como *CocoaZen*, que eran los usuarios sobre los que asentaba la wiki, abandonan el barco.

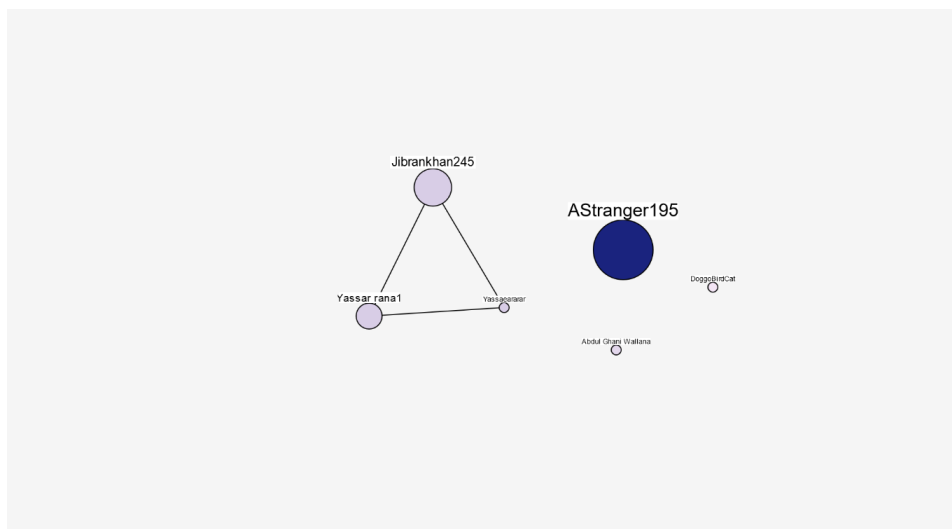


Figura 6.5: Red de coedición entre Nov 2016 - Feb 2017

6.1.3. Conclusiones

Podemos decir que esta era una wiki sustentada sobre dos personas, *CocoaZen* y *Robin Patterson*, las cuales como ya hemos visto actuaban paralelamente para coordinar y revisar el contenido generado por el resto de editores. Podemos confirmar su relevancia si vemos sus páginas de usuario ya que figuran como administradores, [CocoaZen](#), [Robin Patterson](#)

En un principio pensábamos que el contenido de esta wiki iba a fomentar la coordinación de diferentes personas para editar páginas, que en un inicio parece que tiene sentido que tengan una relación entre sí (ciudades de las mismas regiones, o enlaces entre ellas), esto finalmente no lo hemos terminado de confirmar, pues es cierto que hay casos como los del grupo de paquistaníes que sí trabajaban juntos, pero también hay otros tantos que no, como el caso de de ciudades de *EE.UU.* en las que un usuario aislado es el que se encarga de generar contenido para estas.

Es destacable por otro lado, que la política para recibir gente o para atraerla si que funcionara en un principio, ya que al inicio los usuarios principales estaban más dedicados al progreso de esta wiki. El agotamiento de la

comunidad de esta wiki se debe a la falta de usuarios expertos que mantengan un relación estrecha con la wiki, con lo cual, la partida de los que había sumado al hecho de que no haya habido una renovación de estos usuarios "core", fomenta la inactividad de la wiki, lo que se traduce en menos usuarios activos y menos contenido generado.

Podemos confirmar nuestras hipótesis si contrastamos con las series temporales que ofrece *WikiChron-classic* de [la serie total](#). En la figura 6.6 podemos ver la cantidad de nuevos artículos generados en toda la vida de la wiki, y aunque es cierto que no destaca por tener mucha cantidad de artículos y si que podemos ver de manera clara una tendencia hacia la baja. De igual manera, si vemos la figura 6.7 observamos que los usuarios nuevos son cada vez menos, lo que indica que no produce una especial atracción hacia nuevos editores.

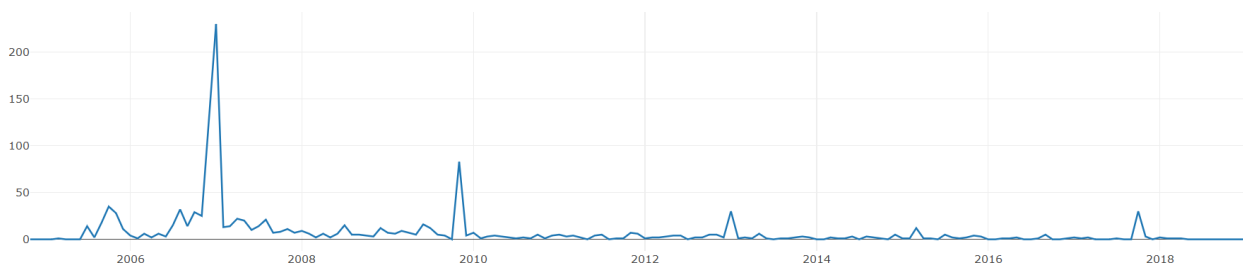


Figura 6.6: Nuevos Artículos



Figura 6.7: Usuarios nuevos

En la tabla 6.2 mostramos un resumen de las métricas más importantes de las redes de co-edición en las etapas citadas. En el capítulo 3.7 se indica lo que implica cada métrica, como interpretarla y en que rangos se mueven los valores.

	Nov 2006 - Feb 2007	Nov 2010 - Feb 2011	Nov 2016 - Feb 2017
Ediciones	890	33	14
Artículos editados	396	16	8
Editores	39	13	6
Enlaces	30	16	3
Densidad	0.04	0.21	0.2
Assortativity	-0.43	1	nan
Gini betweenness	0.94	0.75	nan
Gini degree	0.67	0.54	0.6

Tabla 6.2: Resumen de las métricas más importantes de las redes de coedición

Capítulo 7

Trabajo futuro

Debido a las limitaciones temporales no siempre se pueden implementar todas las ideas que se conciben en un inicio, por lo tanto, planteamos varias mejoras:

- Como mostramos ya en la tabla 3.1, se pueden buscar implementaciones distintas de las redes en base al contenido, por ejemplo, una red bastante útil sería tomar nodos como páginas y que los enlaces sean el número de usuarios que han editado ambas.
- Se pueden incluir diferentes layouts (el actual es CoSE de cytoscape), con el fin de resaltar diferentes partes de la red en base a métricas o comunidades formadas. Referencia [aquí](#).
- Permitir que la gráfica de distribución muestre diferentes métricas, en vez de solo la distribución de grado.
- Añadir la posibilidad de realizar diferentes filtrados a la red, en concreto, filtrar por umbrales (mínimo valor de grado, o peso mínimo de arista). Esta característica es útil para lidiar con redes extensas.
- Añadir la posibilidad de visualizar ego networks (resaltar las conexiones dado un nodo), como se comenta y enseña en Welser et al. (2011), estas nos permiten definir mejor los roles de los usuarios.
- También se podría extraer el núcleo de la app, y lanzar sobre ella una muestra grande de wikis sobre la que realizar un análisis estadístico, con ello se podrían sacar conclusiones más generales acerca de como se comportan estas comunidades.
- Poder añadir redes de proyectos de la *Wikipedia*, pues es tan grande que habría que enfocarlo solo a artículos determinados o categorías concretas.

Capítulo 8

Conclusiones

El trabajo colaborativo permite varios enfoques de estudio, para entender cómo nacen, evolucionan y mueren estas comunidades. El análisis de redes y la modelización de estas comunidades permite un entendimiento más amplio de las dinámicas que mencionamos.

Por todo ello, proponemos una herramienta que permite, no solo el análisis de las comunidades sino también el de los individuos que las forman. Hemos querido seguir la línea de modelización del estado del arte a la hora de tratar las redes, pero a la vez hemos querido incluir la variable temporal que tan importante nos parece a la hora de estudiar las dinámicas que surgen.

Esta herramienta al estar integrada en un mismo ecosistema de *Wiki-Chron*, permite realizar un estudio más amplio, pues puedes disponer de herramientas estadísticas para contrastar los resultados en una misma ventana temporal.

Consideramos que hemos construido un sistema robusto, y de fácil manejo, ya que permite descargar y compartir las redes de una manera ágil y sencilla. Esto fomenta a que si pudiéramos necesitar alguna métrica que la herramienta no provee, nos podamos llevar nuestra red a herramientas generalistas o usarla con librerías especializadas.

Prueba de que el desarrollo ha sido fructífero es haber conseguido escribir un artículo para una conferencia CORE B y específica de colaboración abierta. Todo ello con el poco tiempo del que disponíamos.

Chapter 8

Conclusions

Common-based peer production allows several approaches to study, to understand how these communities are born, evolve and die. Social network analysis provides a broader understanding of the dynamics that we mentioned.

Therefore, we propose a tool that allows, not only the analysis of the communities but also that of the individuals that form them. We have wanted to follow the line of what is already in the state of the art at the moment of modeling the networks, but at the same time we wanted to include the time variable that is so important studying the dynamics that arise.

This tool, being integrated in the same eco-system, allows to carry out a wider study, because you can have statistical tools to compare the results in the same temporal window.

We consider that we have built a robust tool, easy to use, since it allows you to download and share networks in an agile and simple way. This encourages that if we could need some metrics that the tool does not provide, we can take our network and use it with specialized libraries or other tool.

Proof that the development has been successful is to have managed to write a paper with the little time that we had, because this goal has been thanks to the hard work that has involved.

Bibliografía

- ANG, C. S., ZAPHIRIS, P. y PFEIL, U. Cultural Differences in Collaborative Authoring of Wikipedia. *Journal of Computer-Mediated Communication*, vol. 12(1), páginas 88–113, 2006. ISSN 1083-6101.
- ARAZY, O., DAXENBERGER, J., LIFSHITZ-ASSAF, H., NOV, O. y GUREVYCH, I. Turbulent stability of emergent roles: The dualistic nature of self-organizing knowledge coproduction. *Information Systems Research*, vol. 27, páginas 792–812, 2016.
- BARABÁSI, A. y PÓSFAL, M. *Network Science*. Cambridge University Press, 2016. ISBN 9781107076266.
- BOLICI, F., HOWISON, J. y CROWSTON, K. Stigmergic coordination in floss development teams: Integrating explicit and implicit mechanisms. *Cognitive Systems Research*, vol. 38, páginas 14–22, 2016.
- DAXENBERGER, J. y GUREVYCH, I. A corpus-based study of edit categories in featured and non-featured wikipedia articles. *Proceedings of COLING 2012*, páginas 711–726, 2012.
- EL FAQIR, Y., ARROYO, J. y SERRANO, A. Visualization of the evolution of collaboration and communication networks in wikis. *OpenSym19*, 2019.
- FAIGLEY, L. y WITTE, S. Analyzing revision. *College Composition and Communication*, vol. 32(4), páginas 400–414, 1981. ISSN 0010096X.
- FOOTE, J., SHAW, A. y HILL, B. M. Communication networks do not explain the growth or survival of early-stage peer production projects. *SocArXiv*, 2019.
- JIMENEZ-DIAZ, G., SERRANO, A. y ARROYO, J. A wikia census: motives, tools and insights. En *Proceedings of the 14th International Symposium on Open Collaboration (OpenSym'18)*. ACM Press, Article, 6. 2018.

- KADUSHIN, C. *Understanding Social Networks: Theories, Concepts, and Findings*. Oxford University Press, USA, 2012. ISBN 9780195379471.
- KITTUR, A. y KRAUT, R. E. Beyond wikipedia: coordination and conflict in online production groups. En *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, páginas 215–224. ACM, 2010.
- PONS, P. y LATAPY, M. Computing communities in large networks using random walks. En *International symposium on computer and information sciences*, páginas 284–293. Springer, 2005.
- SERRANO, A., ARROYO, J. y HASSAN, S. Participation inequality in wikis: A temporal analysis using wikichron. En *Proceedings of the 14th International Symposium on Open Collaboration*, página 12. ACM, 2018.
- SHAW, A. y HILL, B. M. Laboratories of oligarchy? how the iron law extends to peer production. *Journal of Communication*, vol. 64(2), páginas 215–238, 2014.
- VILLA, D. F. Aplicación de técnicas de visualización y análisis de redes para el estudio de la producción colaborativa de conocimiento. *Director*, vol. 28, página 08, 2018.
- WELSER, H. T., UNDERWOOD, P., COSLEY, D., HANSEN, D. y L, W. B. Wiki networks: Connections of creativity and collaboration. En *Analyzing Social Media Networks with NodeXL* (editado por D. L. Hansen, B. Shneiderman y M. A. Smith), páginas 247 – 271. Morgan Kaufmann, Boston, 2011. ISBN 978-0-12-382229-1.
- ZHANG, X. y WANG, C. Network positions and contributions to online public goods: The case of chinese wikipedia. *Journal of Management Information Systems*, vol. 29(2), páginas 11–40, 2012.